



## Kipnis et al. Respond to "OPEN Questions"

**Victor Kipnis<sup>1</sup>, Amy F. Subar<sup>2</sup>, Arthur Schatzkin<sup>3</sup>, Douglas Midthune<sup>1</sup>, Richard P. Troiano<sup>2</sup>, Dale A. Schoeller<sup>4</sup>, Sheila Bingham<sup>5</sup>, and Laurence S. Freedman<sup>6,7</sup>**

<sup>1</sup> Biometry Research Group, Division of Cancer Prevention, National Cancer Institute, Bethesda, MD.

<sup>2</sup> Applied Research Program, Division of Cancer Control and Population Sciences, National Cancer Institute, Bethesda, MD.

<sup>3</sup> Nutritional Epidemiology Branch, Division of Cancer Epidemiology and Genetics, National Cancer Institute, Bethesda, MD.

<sup>4</sup> Department of Nutritional Sciences, University of Wisconsin, Madison, WI.

<sup>5</sup> Medical Research Council, Dunn Human Nutrition Unit, Cambridge, United Kingdom.

<sup>6</sup> Department of Mathematics, Statistics and Computer Science, Bar Ilan University, Ramat Gan, Israel.

<sup>7</sup> Gertner Institute for Epidemiology and Health Policy Research, Tel Hashomer, Israel.

*Received for publication March 28, 2003; accepted for publication April 3, 2003.*

Abbreviations: DHQ, Diet History Questionnaire; FFQ, food frequency questionnaire; OPEN, Observing Protein and Energy Nutrition; 24HR, 24-hour dietary recall.

We respond briefly to several important and challenging points raised by Dr. Willett (1) concerning our papers published in this issue of the *Journal* (2, 3). First, he notes our use of only 2 days of 24-hour dietary recalls (24HRs) per participant, whereas "usual questionnaire validation studies" (1, p. 22) use many more days. We emphasize that only two independent repeat measurements are required, and in many cases are optimal, for unbiased estimation of within-person variation in dietary report instruments or biomarkers (4).

Second, Dr. Willett's (1) most serious criticism is that we fail to correctly estimate within-person variability in our biomarkers because their repeat measurements were not separated by 6–12 months. He bases his criticism on the analysis of 25 studies by Black and Cole (5) indicating that within-person variation in doubly labeled water measurements increases over time. Because error regarding doubly labeled water includes both physiologic variation and laboratory-specific analytic error, their analysis may be confounded by the potentially substantial between-laboratory variation in analytic error. Moreover, the analyzed studies comprised three types of participants: free living, those for whom activities were enhanced or limited, and those who experienced marked changes in body weight or change in physiologic status (pregnancy or lactation). Because the analyzed studies were based on different numbers of participants and doubly labeled water repeats, the time span between repeats should have been taken into account; the regression should have been fitted by weighted rather than ordinary least squares. With the appropriate anal-

ysis, the slope was actually negative and nonsignificant ( $p = 0.51$ ) in studies that included free-living subjects (the group comparable to those in the Observing Protein and Energy Nutrition (OPEN) Study). Even including the other studies (except those with pregnant/lactating women) did not produce a significant relation between time span and within-person variation ( $p = 0.21$ ). Analysis of urinary nitrogen data (6, 7) with repeated measurements over a period of 1–9 months also showed no statistically significant relation between within-person variation and time span for either women ( $p = 0.72$ ) or men ( $p = 0.42$ ). Therefore, longer-term variability in doubly labeled water and urinary nitrogen does not present a problem, and our estimated within-person variation in both biomarkers is indeed unbiased.

Third, Dr. Willett (1) claims that total energy intake reported on the Diet History Questionnaire (DHQ) is "substantially" lower than that typically estimated by using full-length food frequency questionnaires (FFQs), precluding generalization of OPEN Study results. For women, we showed that median energy estimated by using the DHQ was nearly identical to that of the Block FFQ and 50 kcal lower than that of the Willett FFQ. For men, the median energy estimates from the DHQ were highest among all three FFQs; the Willett FFQ showed an energy value more than 200 kcal lower than the DHQ (8).

Fourth, Dr. Willett states that the data for women are "almost certainly a fluke" (1, p. 23) because the 24HR-based validity of the DHQ for protein density is substantially higher (correlation coefficient, 0.78; standard error, 0.12)

than the typical range of 0.4–0.5 reported in validation studies. In our validation study (8), the DHQ correlation coefficient was 0.61 (standard error, 0.05)—not statistically significantly lower ( $p = 0.29$ ) than that in the OPEN Study. In six European Prospective Investigation into Cancer and Nutrition (EPIC) validation studies with multiple 24HRs as a reference, this correlation for women was 0.44–0.67 (9), and it was 0.84 in the EPIC-Potsdam study (10). Dr. Willett further argues that the confidence interval for the difference in 24HR-based and biomarker-based validity of protein density for women “would have readily included zero” (1, p. 23). Actually, the OPEN Study 24HR overestimated validity by 125 percent, not 60 percent as Dr. Willett claims (Kipnis et al. (3), table 1). The 60 percent figure is true for the estimated attenuation factor. For those who prefer correlation, the difference between a correlation of 0.78 (24HR based) and 0.36 (biomarker based) was statistically significantly different from zero ( $p = 0.03$ ), with a 95 percent confidence interval of 0.11, 0.76.

Finally, Dr. Willett (1) suggests that the OPEN results actually support his previous work on the validity of energy-adjusted nutrients from FFQs because the biomarker-based validity coefficient of the DHQ for energy-adjusted protein is close to the food-record-based coefficient for his FFQ. Dr. Willett argues that, unlike 24HRs, weighed food records do not have correlated errors with FFQs. However, studies with urinary nitrogen suggest that records, no less than recalls, involve both intake-related bias and person-specific bias correlated with that in the FFQ (6, 7). It is plausible that these biases will remain similar in 24HRs and records after energy adjustment.

The fact that Dr. Willett’s (1) validity coefficients are only slightly greater than those in the OPEN Study does not mean that OPEN confirms his results. OPEN results suggest that the actual correlation with true energy-adjusted protein intake for the Willett FFQ in the Nurses’ Health Study may be considerably lower than the record-based estimate of 0.50 (1). Unfortunately, we cannot quantify this coefficient without a biomarker study incorporating records. Therefore, it is extremely important to conduct biomarker studies with

different dietary assessment instruments in prospective cohorts.

## REFERENCES

1. Willett W. Invited commentary: OPEN questions. *Am J Epidemiol* 2003;158:22–24.
2. Subar AF, Kipnis V, Troiano RP, et al. Using intake biomarkers to evaluate the extent of dietary misreporting in a large sample of adults: the OPEN Study. *Am J Epidemiol* 2003;158:1–13.
3. Kipnis V, Subar AF, Midthune D, et al. Structure of dietary measurement error: results of the OPEN biomarker study. *Am J Epidemiol* 2003;158:14–21.
4. Rosner B, Willett WC. Interval estimates for correlation coefficients corrected for within-person variation: implications for study design and hypothesis testing. *Am J Epidemiol* 1988;127:377–86.
5. Black AE, Cole TJ. Within- and between-subject variation in energy expenditure measured by the doubly labelled water technique: implications for validating reported dietary energy intake. *Eur J Clin Nutr* 2000;54:386–94.
6. Kipnis V, Midthune D, Freedman LS, et al. Empirical evidence of correlated biases in dietary assessment instruments and its implications. *Am J Epidemiol* 2001;153:394–403.
7. Kipnis V, Midthune D, Freedman L, et al. Bias in dietary-report instruments and its implications for nutritional epidemiology. *Public Health Nutr* 2002;5:915–23.
8. Subar AF, Thompson FE, Kipnis V, et al. Comparative validation of the Block, Willett, and National Cancer Institute food frequency questionnaires: the Eating at America’s Table Study. *Am J Epidemiol* 2001;154:1089–99.
9. Kaaks R, Slimani N, Riboli E. Pilot phase studies on the accuracy of dietary intake measurements in the EPIC Project: overall evaluation of results. *Int J Epidemiol* 1997;26:S26–36.
10. Kroke A, Klipstein-Grobusch K, Moseneder J, et al. Validation of a self-administered food frequency questionnaire administered in the European Prospective Investigation into Cancer and Nutrition (EPIC) Study: comparison of energy, protein, and macronutrient intakes estimated with the doubly labeled water, urinary nitrogen and repeated 24-h dietary recall methods. *Am J Clin Nutr* 1999;70:439–47.